

Guidelines for the treatment of pneumonia and urinary tract infections: evaluation of methodological quality using the Appraisal of Guidelines, Research and Evaluation II instrument

O. Henig¹, D. Yahav^{2,3}, L. Leibovici^{2,3} and M. Paul^{3,4}

1) Unit of Infectious Diseases and Infection Control, Lady Davis Carmel Medical Centre, Haifa, 2) Department of Medicine E, Rabin Medical Centre, Petah Tikva,

3) Sackler Faculty of Medicine, Tel-Aviv University, Tel-Aviv and 4) Unit of Infectious Diseases, Rambam Health Care Centre, Haifa, Israel

Abstract

Reliance on evidence-based medicine requires high methodological standards from guideline developers. We sought to determine the methodological quality of guidelines on pneumonia and urinary tract infections (UTIs). We included guidelines published by national or international committees in the last 10 years providing recommendations for antibiotic type or duration. We applied the Appraisal of Guidelines for Research and Evaluation II checklist, adding under each item the specific focus relevant to bacterial infections, addressing antibiotic resistance and local epidemiology. Three assessors scored each guideline independently. Mean aggregated scores, converted to percentage per domain, are presented. We included 13 guidelines on the treatment of pneumonia and seven guidelines for the treatment of UTI. 'Scope and purpose' scored 69.4% for pneumonia and 71.4% for UTI. Guidelines were downgraded for lack of an epidemiological overview relevant to intended users. 'Stakeholder involvement' scored 39.5% for pneumonia and 44.5% UTI, with the major fault being lack of patient consultation. 'Rigour of development' scored 42.8% for pneumonia and 56.9% for UTI. Commonly, the search process lacked precision, no risk of bias assessment was performed, outcomes in primary studies were not critically assessed or used to direct recommendations, and there was no formal methodology for formulating recommendations. 'Clarity of presentation' scored highest: 67.7% for pneumonia and 68.5% for UTI. 'Applicability' of the guidelines in antibiotic stewardship programmes was usually not addressed: 16.9% and 25.4%, respectively. 'Editorial independence' scored 30.6% for pneumonia and 55.6% for UTI. Formal examination of guidelines in infectious diseases showed worrying findings related to core methodology and potential bias caused by competing interests.

Keywords: Conflicts of interests, evidence-based medicine, guidelines, recommendations, risk of bias

Article published online: 29 July 2013

Clin Microbiol Infect 2013; **19**: 1106–1114

Corresponding author: M. Paul, Unit of Infectious Diseases, Rambam Health Care Campus, 6 Ha'Aliya Street, PO Box 9602, Haifa 31096, Israel

E-mails: paulm@post.tau.ac.il;
m_paul@rambam.health.gov.il

Background

Globalization and the emphasis on evidence-based medicine have resulted in increasing uniformity between centres worldwide in patient management. This implies increasing reliance on evidence-based guidelines and expert consensus statements rather than basing practice on local habits or opinion. Recommendations in infectious diseases, now more

than ever, have a huge impact on patient management and local policy. This, in turn, may affect local epidemiology.

The responsibility of guidelines developers is therefore to adequately summarize and make use of the available evidence. This includes systematic methods for literature searches, defined criteria for selecting studies, *a priori* definition of the meaningful outcomes to be extracted from the available literature, critical appraisal of the risk of bias, and ranking of

the recommendations by evidence grading. A clear requirement is that the guideline developers be free of conflicts of interests or that these be carefully addressed when they are voting on the final recommendations. These and other methodological requirements from guideline developers have been formally devised and published. However, in infectious diseases, especially with bacterial infections, there are further considerations for guideline development, in addition to the standard criteria relevant to all guidelines in medicine. These pertain to antibiotic resistance. First, the evidence for antibiotic treatment cannot be taken directly from clinical trials, but must be placed in the context of time and location as regards pathogen distribution and antibiotic resistance. Second, the guidelines should consider the ecological impact of treatment recommendations by locale; recommending third-generation cephalosporins universally for hospitalized community-acquired pneumonia (CAP) might lead to the emergence of extended-spectrum β -lactamases in locations where these are rare. The final recommendations should be flexible enough to accommodate different settings worldwide, or otherwise specify the conditions to which the guidelines apply.

Further specific problems in guideline development for bacterial infections include the paucity of evidence. Evidence is often lacking for critically ill patients, who are seldom included in clinical trials. Thus, the patient population included in randomized controlled trials (RCTs) on bacterial infections might not represent the patient population seen in clinical practice [1]. Conceptual trials examining questions such as 'Should we treat this infection? How long should we treat this infection? Should patients be hospitalized?' are rare or non-existent. Trials are usually designed to assess the effects of a specific antibiotic for a specific infection, usually to obtain drug approval. In trials, antibiotics are frequently tested for indications that do not reflect the drugs' use in clinical practice (e.g. levofloxacin and doripenem for urinary tract infection (UTI), and ertapenem for CAP). Guideline developers must use the evidence judiciously to avoid recommendations indirectly led by the industry (e.g. levofloxacin for UTI), and use knowledge and understanding about antibiotics' spectrum of activity, pharmacokinetics and pharmacodynamics to apply clinical trial results to recommended practice.

We sought to determine the methodological quality and reliability of guidelines on bacterial infections, focusing on pneumonia and UTI. We systematically reviewed national and international guidelines on antibiotic treatment of pneumonia and UTI published in the last 10 years. We used the Appraisal of Guidelines for Research and Evaluation II (AGREE-II) criteria to appraise guidelines quality [2], incorporating questions related to bacterial infections within the AGREE-II domains.

Methods

We included guidelines published by national or international committees, organizations or societies on the management of pneumonia and UTI in adults. Within the guidelines, we focused on the recommendations for antibiotic treatment (type and duration). Pneumonia included CAP, hospital-acquired pneumonia, and ventilator-associated pneumonia. UTI included cystitis, pyelonephritis, other complicated UTIs and catheter-associated UTIs in men and women. When more than one version of the guidelines was available, we used the most detailed version (even if available only online). We limited inclusion to guidelines published in English, in the last 10 years (2004–2013).

We searched PubMed, Google, guideline databases, including NICE (<http://guidance.nice.org.uk/CG/Published>) and the Scottish Intercollegiate Guidelines Network (SIGN) (<http://www.sign.ac.uk/guidelines/index.html>), and the websites of professional societies, including the European Society of Clinical Microbiology and Infectious Diseases, the Infectious Diseases Society of America (IDSA), and national thoracic and urological associations. We used the terms 'guidelines', 'recommendations' or 'consensus' crossed with 'pneumonia' and 'urinary tract infection'.

We applied the AGREE-II checklist [3] to each of the guidelines, adding under each item the specific focus relevant to bacterial infections. The checklist comprises 23 items, addressing six domains. Each of the 23 items receives a score between 1 (strongly disagree) to 7 (strongly agree), and the scores of each domain are summarized to give a domain quality score. We did not estimate overall assessment scores. The AGREE-II definitions and the focus with regard to bacterial infections for each item are shown in Table S1. The data pertinent to bacterial infections per domain included the following:

1. Scope and purpose: review of the epidemiology of the disease in the locale targeted by the guidelines. Under 'health question(s) covered by the guideline', we looked for definitions of the outcomes of the infection relevant to patients, using outcome trees or other methods. Under the definitions of the target population, we checked whether age, place of infection acquisition, immunocompromised status and infection severity were addressed, in addition to patient characteristics specific to the types of infections (chronic lung diseases for community-acquired pneumonia; time in hospital for hospital-acquired pneumonia/ventilator-associated pneumonia; sex, catheter and diabetes for UTI).
2. Stakeholder involvement: inclusion of experts from relevant fields in the guideline development group, including

infectious diseases, infection control, microbiology, pharmacology, general practice, geriatrics, intensive care, and acute care, and the relevant society (thoracic societies for pneumonia, and urological/gynaecological societies for UTI); involvement of patients and the public in designing the scope and purpose of the guidelines.

3. Rigour of development: under the item of 'criteria for selecting the evidence', we checked whether inclusion criteria addressed bacterial epidemiology and antibiotic resistance. Under the item of 'strengths and limitations of the body of evidence', we put an emphasis on the AGREE-II-defined criterion of 'appropriateness/relevance of primary and secondary outcomes'. The guidelines should have addressed the effect of different antibiotic choices on the outcome(s) of relevance, as identified under 'scope and purpose'. Under the item 'health benefits, side effects, and risks', we checked whether guidelines addressed the ecological impact of the antibiotic recommendations.
4. Clarity of presentation: we checked whether antibiotic treatment recommendations were stratified at least by place of infection acquisition, age and infection severity; and whether the recommendations addressed the epidemiological settings for which the recommendations are relevant, or recommendations were stratified by epidemiological setting. Under 'different options for management of the condition or health issue are clearly presented', we checked whether alternatives that have not been tested in RCTs, with a similar spectrum of coverage and activity, were addressed.
5. Applicability: we examined whether the guidelines addressed antibiotic stewardship programmes and how to implement recommendations within such programmes; provided directions for antibiotic audits; and addressed performance measures relevant to the guidelines.
6. Editorial independence: as defined by AGREE-II.

Before the rating process was begun, the contents of each item of the AGREE-II checklist were discussed (Table S1), scoring was piloted by three assessors, and differences were discussed. The three assessors then applied the checklist to each guideline independently, detailing the reasons for the scoring of each item (available from the authors on request). Large differences (>4 points) between assessors were discussed, and omissions or errors were corrected, although no attempt was made to reach a consensus. The final scores of the three reviewers were aggregated and converted to a percentage for each domain, as recommended [4]. The correlation of the scores between pairs of assessors was examined by use of Spearman's correlation coefficient.

Pooled mean/median scores for all guidelines in each category (pneumonia and UTI) were calculated for each domain.

Results

Our search resulted in 2053 references, among which 42 potentially relevant guidelines for pneumonia or UTI treatment issued by formal national or international associations were identified. Twenty-two were excluded because of overlap with another included guideline or non-English language, and 13 guidelines on the treatment of pneumonia [5–17] and seven guidelines for the treatment of UTI [18–24] were included. The guidelines were published between 2004 and 2013, and spanned all the subtypes of infection under each diagnosis. The guidelines originated from Europe, the USA, Asia, Brazil, South Africa, Japan, and the Gulf Cooperation Council. The correlations of the scores between the three pairs of assessors for all items and all guidelines were 0.828, 0.815, and 0.746 ($p < 0.001$), indicating high correlation. Aggregated results of the AGREE-II scores and pooled results of all guidelines are provided in Tables 1 and 2. Detailed scores per guideline are provided in Tables S2 and S3. The top-ranking antibiotic recommendations per guidelines are shown in Tables S4 and S5.

Scope and purpose

This item is concerned with 'the overall aim of the guideline, the specific health questions, and the target population'. The means of the aggregated scores were 69.4% (median, 72.2%; range, 48.2–87%) for pneumonia, and 71.4% (median, 74.1%; range, 33.3–87%) for UTI.

Although the objectives of the guidelines (item 1) were generally described, nearly all guidelines were downgraded for lack of statements on the expected health benefits or impact of the guidelines on society, populations of patients, or individuals. As many of the guidelines were updates of previous versions, we searched for some assessment of the impact of the previous version of the guidelines, but did not find such assessments. Under description of the health question (item 2), eight of 13 guidelines on pneumonia and only two of seven guidelines on UTI defined and described the outcomes relevant to the infection. Most guidelines presented an epidemiological overview of the causative bacteria and their resistance profiles in general, but not all reviewed data specifically relevant to the settings targeted by the guidelines. Of the pneumonia guidelines, six of 13 guidelines addressed the epidemiology relevant to the targeted users; of the UTI guidelines, two of seven provided

TABLE 1. Appraisal of Guidelines for Research and Evaluation II scoring of guidelines for the treatment of pneumonia

Issuing society	Year of publication	Topic(s) addressed by the guidelines	Scope and purpose (%)	Stakeholder involvement (%)	Rigour of development (%)	Clarity of presentation (%)	Applicability (%)	Editorial independence (%)
Swedish Society of Infectious Diseases [5]	2012	CAP	53.7	20.4	36.8	75.9	1.4	94.4
Dutch Working Party on Antibiotic Policy/Dutch Association of Chest Physicians [6]	2012	CAP	79.6	44.4	65.3	92.6	25	25.0
European Society of Clinical Microbiology and Infectious Diseases/European Respiratory Society [7]	2011	LRTI	63.0	51.9	50	70.4	0.0	25.0
Spanish Society of Pulmonology and Thoracic Surgery [8]	2010	CAP	64.8	0.0	20.8	55.6	1.4	0.0
British Thoracic Society [14]	2009	CAP	83.3	66.7	80.6	96.3	48.6	44.4
Brazilian Thoracic Association [9]	2009	CAP	46.3	7.4	25.7	38.9	2.8	44.4
American Burn Association [10]	2009	VAP	72.2	38.9	39.6	33.3	2.8	0.0
British Society of Antimicrobial Chemotherapy [11]	2008	HAP	74.1	55.6	66.7	66.7	13.9	36.1
South African Thoracic Society [12]	2007	CAP	83.3	25.9	23.6	57.4	4.2	0.0
Gulf Cooperation Council ^a [13]	2007	CAP	66.7	70.4	31.9	74.1	37.5	0.0
Infectious Diseases Society of America/American Thoracic Society [15]	2007	CAP	87.0	68.5	54.9	92.6	43.1	58.3
Infectious Diseases Society of America/American Thoracic Society [16]	2005	HAP/VAP/HCAP	87.0	35.2	41.6	81.5	36.1	69.4
Japanese Respiratory Society ^b [17]	2004	HCAP	48.2	27.8	18.8	44.4	2.8	0.0
Average			69.4	39.5	42.8	67.7	16.9	30.6
SD			14.3	22.9	19.6	20.8	18.5	30.96
Median			72.2	38.9	39.6	70.4	4.2	25.0

CAP, community-acquired pneumonia; HAP, hospital-acquired pneumonia; HCAP, healthcare-associated pneumonia; LRTI, lower respiratory tract infection; SD, standard deviation; VAP, ventilator-associated pneumonia.

^aCooperative framework of six countries including the United Arab Emirates, the Kingdom of Bahrain, the Kingdom of Saudi Arabia, the Sultanate of Oman, the State of Qatar, and the State of Kuwait.

^bThe Japanese Respiratory Society guidelines were published in several papers that were included in the review. The full reference list is shown in Tables S1–S5.

locally relevant bacteriological data. Population description (item 3) scores were downgraded mainly for lack of definitions of patient subgroups relevant to the studied infection in advance.

Stakeholder involvement

This item focuses on ‘the extent to which the guideline was developed by the appropriate stakeholders and represents the views of its intended users’. The means of the aggregated scores were very low, at 39.5% (median, 38.9%; range, 0–70.4%) for pneumonia, and 44.5% (median, 37%; range, 5.6–100%) for UTI.

The scores were low for all three items included under this domain. The composition of the guideline development group or authorship team (item 4) received the highest score of the three, but was poorly adapted to the topic or poorly described in most guidelines. The guideline panel of the high-scored guidelines [6,11,13,14,18–20,22] represented experts from microbiology, infectious diseases, pharmacology, medical and intensive care, general practice, geriatrics, and public health, and the relevant specialists (pulmonology, thoracic diseases, urology, and gynaecology). Of the authors listed in UTI guidelines, 29% were women (45% in the SIGN guidelines [19] and 24% in all others). The lowest score was given to seeking

the views and preferences of the target population (item 5). All but three guidelines were scored 1 by all assessors (strongly disagree). Two guidelines [11, 19] were subjected to public consultation before publication, and the IDSA guidelines for CAP [15] addressed consumer preference, but only for quality indicators. However, none of the guidelines described a process of explicit patient consultation. The target users of the guidelines (item 6) were generally not specifically described, but all assessors felt that this item is probably less relevant to our topic, as target users are defined by the scope of the guidelines.

Rigour of development

This domain relates to ‘the process used to gather and synthesize the evidence, and the methods used to formulate the recommendations and to update them’. It is a large domain, including eight items, and is the core of the guideline methodology. The mean scores were 42.8% (median, 39.6%; range, 18.8–66.7%) for pneumonia, and 56.9% (median, 58.3%; range, 23.6–89.6%) for UTI.

The search process (items 7 and 8) scored low both for pneumonia and for UTI. No search strategy was provided or referred to in seven of 13 guidelines for pneumonia and three of eight guidelines for UTI (item 7). Otherwise, scoring points

TABLE 2. AGREE-II scoring of guidelines for the treatment of UTI

Issuing society	Year of publication	Topic(s) addressed by the guidelines	Scope and purpose (%)	Stakeholder involvement (%)	Rigour of development (%)	Clarity of presentation (%)	Applicability (%)	Editorial independence (%)
The Dutch Working Party on Antibiotic Policy [18]	2013	Complicated UTI, CA-UTI	87	37	58.3	81.5	31.9	25
Scottish Intercollegiate Guidelines Network [19]	2012	Cystitis ^a , pyelonephritis, complicated UTI, CA-UTI	74.1	100	76.4	61.1	68.1	47.2
European Association of Urology [21]	2011	Cystitis, pyelonephritis, complicated UTI, CA-UTI, urological infections	72.2	35.2	39.6	88.9	0	66.7
Infectious Diseases Society of America/European Society of Clinical Microbiology and Infectious Diseases [20]	2011	Cystitis, pyelonephritis	87	31.5	89.6	94.4	29.2	80.6
Infectious Diseases Society of America [22]	2009	CA-UTI	77.8	59.3	77.1	55.6	38.9	75
American College of Obstetrics and Gynecology [23]	2008	Cystitis, pyelonephritis	68.5	5.6	23.6	50	9.7	0
European/Asian [24]	2008	CA-UTI	33.3	42.6	34	48.2	0	94.4
Average			71.4	44.5	56.9	68.5	25.4	55.6
SD			18.2	29.2	25.1	19.3	24.5	33.5
Median			74.1	37	58.3	61.1	29.2	66.7

CA-UTI, catheter-associated urinary tract infection; SD, standard deviation; UTI, urinary tract infection.

^aRelevant for women only.^bEuropean Association of Urology, European Society for Infection in Urology, Urological Association of Asia, the Asian Association of Urology/STD, the Western Pacific Society for Chemotherapy, the Federation of European Societies for Chemotherapy and Infection, and the International Society of Chemotherapy for Infection and Cancer.

were deducted for a restricted (only PubMed), non-structured search. Interventions were not included in the search strategies of all guidelines. The criteria for selecting the evidence (item 8) were scored even lower than the methods of the search, as these were not usually defined. When they were defined, the selection criteria did not include epidemiological considerations that would result in a set of studies relevant with regard to bacterial distribution and antibiotic resistance.

An intention to critically appraise the evidence (item 9), as reflected by using a grading table, was present in all but two guidelines for pneumonia. However, evidence grading was limited to the classification of studies by design. None of the guidelines performed a risk of bias assessment of individual trials, addressing selection, performance, detection, attrition, reporting or other biases, except for the SIGN guidelines for UTI [19]. None of the guidelines used the GRADE methodology. More importantly, in our view, was that almost none of the guidelines included an assessment of the relevance of the outcomes of individual studies, or a statement on the appropriateness of the antibiotics compared. The IDSA/American Thoracic Society guidelines on CAP were an exception with regard to outcome definition [15], defining mortality as the relevant outcome, and selecting studies for inclusion by using this outcome. The methods for formulating the recommendations (item 10) were scored very low for both pneumonia and UTI guidelines, as almost none of the guidelines described a formal process of voting or other techniques to reach a consensus on specific recommendations. A few guidelines described an informal process of meetings, emails, review of drafts, and discussions, and received an intermediate score. High scores were given to single guidelines using Delphi methods to combine expert opinions, a voting process to define recommendations, or the 'SIGN' methodology, which is a formal process based on grading scores [14,15,19].

Considering the health benefits, side effects and risks in formulating the recommendations (item 11) received a medium score for both pneumonia and UTI. Surprisingly, in guidelines, there was usually a dissociation between outcomes and recommendations for treatment. The reasoning provided for antibiotic selection was their expected spectra of coverage. Clinical trials were sometimes (infrequently) quoted in favour of specific recommendations, but the outcomes assessed in the trials and on which the guideline developers based their recommendations were not discussed. Similarly, the ecological implications of the different antibiotics were not explicitly discussed. Often, guidelines provided a list of antibiotic options for a certain type of infection, mainly for hospital-acquired infections, as appropriate. Although, as previously noted, most guidelines defined an evidence grading score, surprisingly, in approximately half of the guidelines, the final recommenda-

tions were not linked to an evidence grade (item 12, explicit linking between the recommendations and the supporting evidence). Similarly, approximately half of all guidelines did not provide references to primary studies that support the recommendations.

External review of the guidelines (item 13) was specifically reported in few guidelines: three of 13 for pneumonia, and four of seven for UTI. Naturally, guidelines published in peer-reviewed journals were reviewed; however, a true external review should be part of the guideline development process prior to submission of the final version for publication. Finally, in this domain, only a few guidelines (two for pneumonia and two for UTI) noted a plan for future updates. We gave a low, but not minimal, score to guidelines that were updates of previous versions.

Clarity of presentation

This domain deals with 'the language, structure and format of the guideline'. Relative to other domains, this domain scored high overall, with a mean of 67.7% (median, 70.4%; range, 33.3–96.3%) for pneumonia, and a mean of 68.5% (median, 61.1%; range, 48.2–94.4%) for UTI.

Clarity and avoiding ambiguity (item 15) was scored high in most guidelines. Guidelines usually stratified recommendations by clinically relevant subgroups. The highest scores were given to guidelines that also addressed the epidemiological settings for which the recommendations are relevant. Guidelines for hospital-acquired infections did not always refer to specific antibiotics. However, all assessors considered statements such as 'antibiotic choice should be directed by the local epidemiology' as appropriate, provided that some directives were given on how to use local epidemiology and on duration of treatment. Within this domain, the lowest scores were given to item 16: 'clear presentation of the different options for management of the condition or health issue'. Guidelines were downgraded when they provided a list of antibiotic options with no grading, ranking or explanation of how to choose between the options. Most did not consider alternatives to antibiotics tested in RCTs. Finally, high scores were given to most guidelines for item 17: 'key recommendations are easily identifiable'. Recommendations were usually presented in tables, figures, or summary statements.

Applicability

The applicability domain pertains to 'the likely barriers and facilitators to implementation, strategies to improve uptake, and resource implications of applying the guideline'. This domain received a very low score, as most guidelines did not address these items: a mean of 16.9% (median, 4.2%; range, 0–

48.6%) for pneumonia, and a mean of 25.4% (median, 29.2%; range, 0–68.1%) for UTI.

None of the guidelines analysed facilitators of and barriers to guideline application (item 18). The higher-scored guidelines provided some advice or tools for putting the recommendations into practice (item 19). These were in the form of recommendations on how to adapt the guidelines locally, who should undertake implementation and dissemination, recommendations for implementation, and dissemination strategies [13,19]. Monitoring and/or auditing criteria (item 21) were well described in few guidelines, which proposed performance measures [6,15] or a formal audit tool [14] for CAP, and auditing measures for UTI [19]. The American Congress of Obstetricians and Gynecologists guidelines for UTI suggested a single performance measure: the percentage of women with pyelonephritis treated for 14 days [23]. This guideline was scored low, as the evidence backing this measure was lacking.

Nearly none of the guidelines considered the potential resource implications of applying the recommendations (item 20): ten of 13 guidelines for pneumonia and five of seven guidelines for UTI were given the minimal score by all three assessors.

Editorial independence

This domain is concerned with processes and appropriate statements ensuring that the formulation of recommendations was not unduly biased by competing interests. The scores were low, with a mean of 30.6% (median, 25%; range, 0–94.4%) for pneumonia, and a mean of 55.6% (median, 66.7%; range, 0–94.4%) for UTI.

The higher-scored guidelines explicitly reported funding (item 22) by a non-profit organization, referred to the funding of all stages of guideline development and meetings of the authors, and declared no conflicts of interests (item 23) for all authors. The guidelines detailing all conflicts of interests of the authors never provided an explanation of how these were dealt with (e.g. exclusion of the author when voting on an intervention with a possible conflict of interest). These guidelines were given an intermediate score.

Discussion

Guidelines on the treatment of pneumonia and UTI scored reasonably well with respect to scope and purpose definitions and the clarity of presentation of the final recommendations. All other domains were scored dismally low, with a pooled score below 50% for all domains in pneumonia and most domains in UTI. These domains refer to the core methodology of the guideline's development process, the quality of the

actual recommendations, patient involvement, guideline implementation, and conflicts of interests. Guidelines published by professional guideline development groups, most prominently the SIGN guidelines, scored highest on methodology. However, they scored low for parameters specifically relevant to bacterial infections, such as epidemiology and ecological considerations. Guidelines developed by societies of infection and microbiology scored low on methodology, and, surprisingly, most also scored low for the epidemiological considerations.

Formal examination of the guidelines with the AGREE-II score revealed two important limitations of current guidelines in infectious diseases: interventions and outcomes. First, the interventions (i.e. antibiotic considered to be relevant for the infection) were not defined in advance (no definition, and no structured search strategy). When the evidence was reviewed, there was no critical appraisal of the relevance of the antibiotics identified. Guidelines should start with clinical questions: population, well-defined interventions, and main outcomes of interest. A natural process would be to define in advance the antibiotics considered to be relevant for the infection and search for these. Given the common coverage and mechanism of action of different antibiotics, one could consider defining in advance antibiotics that will be used as indirect evidence for the efficacy of another antibiotic, e.g. inferring on the efficacy of ciprofloxacin from trials assessing levofloxacin. With such a process, trials of ertapenem for CAP would probably not be included *a priori*. The strategy used by most current guidelines is to sum up phase 3/4 non-inferiority trials conducted as part of the drug approval process, with little criticism on the relevance of the antibiotics compared.

Second, outcomes were usually not defined in advance, the relevance of the outcomes examined in the primary studies were not discussed, and, most interestingly, the relationships between actual recommendations and outcomes in primary studies were rarely described. None of the guidelines addressed the outcomes targeted when giving recommendations for antibiotic treatment. The guidelines addressed the coverage obtained with the different antibiotic choices, but almost none addressed clinical outcomes in the antibiotic recommendation section (i.e. why is it important to achieve adequate coverage and what we want to achieve by treating patients; an exception was noted for the IDSA/ESCMID guidelines for cystitis [20]). The patient perspective was never addressed. The result is dissociation between what we expect from antibiotic treatment in clinical practice and the recommendations.

A structured search and risk of bias assessment is routine in systematic reviews. Risk of bias assessment is incorporated in the evidence-grading methods using GRADE methodology [25]. Guidelines have advanced in the last decade from

non-evidence-based expert opinion to a semi-systematic process with grading of evidence by study design. However, there is a still large gap between this and a truly systematic process. Currently, the methods for evidence grading in all of the guidelines that we identified were developed internally, were variable, were non-comparable, and were based only on study design. Most grading tables had important limitations. For example, the IDSA evidence-grading system does not address systematic reviews, and, consequently, systematic reviews are conspicuously missing from the evidence review in most recommendations. The AGREE-II and PRISMA statements [3,26] recommend that the search strategy and selection criteria address population, interventions, comparisons, and outcomes (PICO). None of the available guidelines formulated such a strategy. Admittedly, all guidelines had a broader scope than treatment alone, and broad searches of the field were therefore conducted. However, multiple search strategies should optimally be designed for each of the questions addressed by the guidelines. The search strategy for antibiotic treatment should address the four PICO components. The overall systematic process should ensure recommendations using reasonable antibiotics and relying on patient-relevant outcomes.

We searched for epidemiological and ecological considerations in the formulation of the guidelines. We found inconsistent matching of antibiotic treatment recommendations to local bacterial distribution and antibiotic resistance. The ecological considerations of selecting one or another antibiotic from a list of options, in different epidemiological settings, were rarely discussed. This is important to ensure adequate coverage of the recommended antibiotics, but also to avoid recommending antibiotics that are too broad-spectrum for locations with low baseline resistance. A study examining the predicted changes in antibiotic use in The Netherlands estimated that there would be a large increase in antibiotic use if the IDSA guidelines for CAP were to be implemented there, with potential important ecological consequences [27].

Patient consultation is relevant for any disease. For UTI, especially uncomplicated UTI among young women, consulting healthy women and patients is particularly relevant to direct management. Studies have shown that women frequently agree to defer antibiotic treatment, and in most of them symptoms resolve spontaneously [28, 29]. We did not identify an explicit process of patient consultation in all guidelines.

A very low score was given to the methods for formulating the recommendations and the domain of editorial independence. A formal process to decide on the final recommendations is important to limit inappropriate influences. Conflicts of interests are common and almost unavoidable in guideline

panels. Not only should conflicts of interests be reported, but it is very important that users of guidelines be aware of how conflicts were dealt with (e.g. exclusion of a person with a conflict from contributing to a specific recommendation). Finally, recommendations for treatment will only have an effect if effectively implemented. Antibiotic stewardship and guideline application is a major issue in infectious disease. Existing guidelines rarely addressed the implementation of recommendations within antibiotic stewardship programmes.

Our analysis is limited by the use of a panel of assessors who were all from the field of infectious diseases and who were not blinded to the source of the guidelines. We excluded guidelines that were not published in English, as we targeted guidelines that might be read and used worldwide. It is possible that guidelines published in local languages perform better with respect to consideration of local epidemiology. We did not estimate overall assessment scores, also recommended by AGREE-II. We opted for the more objective and defined scoring of individual items. Our final scores were lower than those reported in a systematic review examining AGREE-II scores for WHO guidelines after the establishment of a Guidelines Review Committee [30]. These guidelines addressed treatment, prevention and diagnosis in the fields of human immunodeficiency virus, tuberculosis, malaria, influenza, and hand hygiene. Our scores might have been lowered by the addition of content relevant to bacterial infections to the score items. We are not aware of other evaluations of guidelines on bacterial infections using the AGREE-II score.

In conclusion, we found wide room for improvement in guidelines on pneumonia and UTI. Specifically, systematic methods for formulating the questions addressed by the guidelines should result in better selection of antibiotics targeting patient-relevant outcomes. A unique feature of guidelines in infectious diseases is the need to address local epidemiology.

Transparency Declaration

The authors declare no conflicts of interest.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1. AGREE-II score items and interpretation specific to bacterial infections.

Table S2. Scores of the 23 AGREE-II items for guidelines on pneumonia.

Table S3. Scores of the 23 AGREE-II items for guidelines on UTI.

Table S4. Top-ranking antibiotic(s) recommendations for pneumonia.

Table S5. Top-ranking antibiotic(s) recommendations for UTI.

References

1. Falagas ME, Vouloumanou EK, Sgouros K, Athanasiou S, Peppas G, Siempos II. Patients included in randomised controlled trials do not represent those seen in clinical practice: focus on antimicrobial agents. *Int J Antimicrob Agents* 2010; 36: 1–13.
2. Burgers JS, Fervers B, Haugh M et al. International assessment of the quality of clinical practice guidelines in oncology using the Appraisal of Guidelines and Research and Evaluation instrument. *J Clin Oncol* 2004; 22: 2000–2007.
3. Brouwers MC, Kho ME, Browman GP et al. AGREE II: advancing guideline development, reporting and evaluation in health care. *CMAJ* 2010; 182: E839–E842.
4. The AGREE Next Steps Consortium: APPRAISAL OF GUIDELINES FOR RESEARCH & EVALUATION II—User's manual and AGREE-II instrument. Available at: http://www.agreetrust.org/wp-content/uploads/2013/06/AGREE_II_Users_Manual_and_23-item_Instrument_ENGLISH.pdf (last accessed 9 August 2013).
5. Spindler C, Stralin K, Eriksson L et al. Swedish guidelines on the management of community-acquired pneumonia in immunocompetent adults—Swedish Society of Infectious Diseases 2012. *Scand J Infect Dis* 2012; 44: 885–902.
6. Wiersinga WJ, Bonten MJ, Boersma WVG et al. SWAB/NVALT (Dutch Working Party on Antibiotic Policy and Dutch Association of Chest Physicians) guidelines on the management of community-acquired pneumonia in adults. *Neth J Med* 2012; 70: 90–101.
7. Woodhead M, Blasi F, Ewig S et al. Guidelines for the management of adult lower respiratory tract infections—full version. *Clin Microbiol Infect* 2011; 17 (suppl 6): E1–E59.
8. Menendez R, Torres A, Aspa J, Capelastegui A, Prat C, Rodriguez de Castro F. Community acquired pneumonia. New guidelines of the Spanish Society of Chest Diseases and Thoracic Surgery (SEPAR). *Arch Bronconeumol* 2010; 46: 543–558.
9. Correa Rde A, Lundgren FL, Pereira-Silva JL et al. Brazilian guidelines for community-acquired pneumonia in immunocompetent adults—2009. *J Bras Pneumol* 2009; 35: 574–601.
10. Mosier MJ, Pham TN. American Burn Association Practice guidelines for prevention, diagnosis, and treatment of ventilator-associated pneumonia (VAP) in burn patients. *J Burn Care Res* 2009; 30: 910–928.
11. Masterton RG, Galloway A, French G et al. Guidelines for the management of hospital-acquired pneumonia in the UK: report of the working party on hospital-acquired pneumonia of the British Society for Antimicrobial Chemotherapy. *J Antimicrob Chemother* 2008; 62: 5–34.
12. Feldman C, Brink AJ, Richards GA, Maartens G, Bateman ED for the Working Group of the South African Thoracic Society. Management of community-acquired pneumonia in adults. *S Afr Med J* 2007; 97: 1296–1306.
13. Memish ZA, Arabi YM, Ahmed QA, Shibl AM, Niederman MS. Executive summary of the Gulf Cooperation Council practice guidelines for the management of community-acquired pneumonia. *J Chemother* 2007; 19 (suppl 1): 7–11.

14. Lim WS, Baudouin SV, George RC *et al.* BTS guidelines for the management of community acquired pneumonia in adults: update 2009. *Thorax* 2009; 64 (suppl 3): iii1–iii55.
15. Mandell LA, Wunderink RG, Anzueto A *et al.* Infectious Diseases Society of America/American Thoracic Society consensus guidelines on the management of community-acquired pneumonia in adults. *Clin Infect Dis* 2007; 44 (suppl 2): S27–S72.
16. American Thoracic Society; Infectious Diseases Society of America. Guidelines for the management of adults with hospital-acquired, ventilator-associated, and healthcare-associated pneumonia. *Am J Respir Crit Care Med* 2005; 171: 388–416.
17. The committee for The Japanese Respiratory Society guidelines in management of respiratory infections. Principles for the development of the guidelines. *Respirology* 2004; 9 (suppl 1): S1–S2.
18. Dutch Working Party on Antibiotic Policy (SWAB): SWAB guidelines for antimicrobial therapy of complicated urinary tract infections in adults. Available at: <http://www.swab.nl/richtlijnen> (last accessed 9 August 2013).
19. Scottish Intercollegiate Guidelines Network (SIGN). *Management of suspected bacterial urinary tract infection in adults*. Edinburgh: SIGN, 2012. SIGN publication no. 88. Available at: <http://www.sign.ac.uk> (last accessed July 2012).
20. Gupta K, Hooton TM, Naber KG *et al.* International clinical practice guidelines for the treatment of acute uncomplicated cystitis and pyelonephritis in women: a 2010 update by the Infectious Diseases Society of America and the European Society for Microbiology and Infectious Diseases. *Clin Infect Dis* 2011; 52: e103–e120.
21. Grabe M, Bjerklund-Johansen TE, Botto H *et al.* Guidelines on urological infections. Available at: <http://www.uroweb.org/guidelines/online-guidelines/> (last accessed 9 August 2013).
22. Hooton TM, Bradley SF, Cardenas DD *et al.* Diagnosis, prevention, and treatment of catheter-associated urinary tract infection in adults: 2009 International Clinical Practice Guidelines from the Infectious Diseases Society of America. *Clin Infect Dis* 2010; 50: 625–663.
23. American College of Obstetrics and Gynecologists. ACOG Practice Bulletin No. 91: treatment of urinary tract infections in nonpregnant women. *Obstet Gynecol* 2008; 111: 785–794.
24. Tenke P, Kovacs B, Bjerklund Johansen TE, Matsumoto T, Tambyah PA, Naber KG. European and Asian guidelines on management and prevention of catheter-associated urinary tract infections. *Int J Antimicrob Agents* 2008; 31 (suppl 1): S68–S78.
25. Guyatt GH, Oxman AD, Vist GE *et al.* GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008; 336: 924–926.
26. Liberati A, Altman DG, Tetzlaff J *et al.* The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ* 2009; 339: b2700.
27. Oosterheert JJ, Bonten MJ, Schneider MM, Hoepelman IM. Predicted effects on antibiotic use following the introduction of British or North American guidelines for community-acquired pneumonia in The Netherlands. *Clin Microbiol Infect* 2005; 11: 992–998.
28. Arav-Boger R, Leibovici L, Danon YL. Urinary tract infections with low and high colony counts in young women. Spontaneous remission and single-dose vs multiple-day treatment. *Arch Intern Med* 1994; 154: 300–304.
29. Knottnerus BJ, Geerlings SE, Moll van Charante EP, Ter Riet G. Women with symptoms of uncomplicated urinary tract infection are often willing to delay antibiotic treatment: a prospective cohort study. *BMC Fam Pract* 2013; 14: 71.
30. Sinclair D, Isba R, Kredo T, Zani B, Smith H, Garner P. World Health Organization guideline development: an evaluation. *PLoS ONE* 2013; 8: e63715.